SAS GLOBAL FORUM 2020 #SASGF

Paper 4853-2020

Multinomial vs. Ordinal. Does model selection make a difference? for SAS® Global Forum 2020

Corey Leadbeater, 8People, Inc. and National University

ABSTRACT

Regression modeling, a foundational component of data analysis and machine learning, is one of the most highly sought-after skills by employers seeking new data scientists.[1] While most data science curricula tend to include regression modeling techniques, the conceptual nuances between theoretical and practical applications can be nebulous. In this paper we use SAS 9.4 and the 2016 Monitoring Futures Survey to demonstrate the utility and effects of model selection on context. That is, the ability of the model to properly communicate the users intended information.

A cross-sectional secondary analysis using Pearson's Chi-Square test statistic for independence was conducted on the 2016 Monitoring Futures Study dataset to determine the association between "behavior risk" (v7335) and each of the three predictor variables, "parental communication" (v7254), "time spent alone after school" (loner) and student letter grades (v7221). Next, two logistic models were computer using these predictors. Statistically significant associations were identified, and adjusted odds ratios were produced using both multinomial and ordinal regression models. Output from both models were evaluated and compared to demonstrate the utility of ordinal modeling (and output) as a "higher-view" and generalizing procedure, whereas multinomial models are more appropriate for a more detailed view of group-level comparisons.

INTRODUCTION

Regression modeling, a foundational component of data analysis and machine learning, is one of the most highly sought-after skills by employers seeking new data scientists.[1] While most data science curricula tend to include regression modeling techniques, the conceptual nuances between theoretical and practical applications can be nebulous. In this paper we use SAS 9.4 and the 2016 Monitoring Futures Survey to demonstrate the utility and effects of model selection on context. That is, the ability of the model to properly communicate intended information.

Our case study explores the need for a behavioral intervention program for students in the eight and tenth grades. For programs like these, the investigators choice of model selection is a critical part of ensuring appropriate program enrollment. Here we provide a case study to illustrate the contextual effects of multinomial vs. ordinal model selection.

A cross-sectional secondary analysis using Pearson's Chi-Square test statistic for independence was conducted to determine the association between our dependent variable, "behavior risk" (v7335), and each of the three predictor variables, "parental communication" (v7254), "time spent alone after school" (loner) and student letter grades (v7221). Two logistic models, multinomial and ordinal, were then generated and statistically significant associations were identified in both cases. Model outputs were compared to demonstrate the utility of ordinal modeling as a generalizing procedure while multinomial modeling proved the better selection for detailing specific group-level comparisons.

STUDY POPULATION AND DATA

The study was performed using data from the 2016 Monitoring Futures Study eighth and tenth grade samples and contained 30,858 observations after controlling for refused or missing answers across all variables. Missing or refused variables were recoded to “.” for respective categories. The analysis and logistic regressions used 4 variables, which were all found to be statistically significant ($p < .0001$) using Pearson’s Chi-Square statistic. Variables include:

V7335 – “school behavior risk”

V7254 – “level of parental communication”

Loner – “amount of time (in hours) spent alone after school”

Grades – “letter grades of student” to measure academic performance

Student behavior risk was rated on a 1 – 5 Likert scale indicating the number of times a student is sent to the office due to poor behavior (1 = never, 2 = rarely, 3 = sometimes, 4 = often 5 = always) , level of parental communication on a 1 – 3 Likert scale (1 = none, 2 = low, 3 = high). The variable V7221 was recoded to a 4 tiered categorical variable “grades” (1 = D, 2 = C, 3 = B, 4 = A) and variable v7214 was recoded to “Loner” and measured on a 1 – 3 Likert scale (1 = not alone, 2 = 1-3 hours alone, and 3 = 4 or more hours alone per day after school).

Descriptive and Univariate statistics were collected to evaluate skewness, kurtosis, distribution, amount of missing data or other potential issues with distribution of all variables. To evaluate our missing data a “do-loop” was invoked against a specially created array. The code is shown below:

```
Data check ;
  set mtf ;
  array chckmiss{*} v501 v7221 v7202 v7335 v7254 loner grade ;
  missdata = 0 ;
  do i = 1 to dim(chckmiss) ; /* dim(chckmiss) = set dimensions
  to however variables are in chckmiss */
     if chckmiss{i} = . then missdata = missdata + 1 ;
  end ;
  if missdata > 0 then anymiss = 1 ;
     else anymiss = 0 ;
run;
*** Check for missing data differences by grade ;
proc freq data =check ;
  tables anymiss missdata ;
  run ;
proc freq data=check ;
  tables   anymiss*(v501 v7221 v7202 v7335 v7254 loner grade) ;
  run ;
  proc ttest data=check ;
  class v501 ;
    var missdata ;
    run;
```

Table 1, below, shows the variables to have minimal missing values and mostly normal distributions, except for v7254 (Talking with parents about problems), which was missing data for about 1/3 of students. A t-test (Table 2) was conducted by grade with number of missing items (‘missdata’). This was significant, with an N of 32,873. The distributions in Figure 1 show similarities across the two grades. The mean number of items missing for eight-graders was 0.55 and for tenth-graders was 0.46. Interestingly, the mode for grades was 9=(A), which calls into question the validity of this self-reported grade data, however, several possibilities exist. For example, higher performing students could be expected to participate in the study in greater numbers than lower performers. Additionally, participating schools may have willingly engaged their highest performing students resulting in selection a bias.

Table 1. Descriptive statistics for numeric variables

Descriptive statistics for numeric variables

The MEANS Procedure

Variable	Label	N	Mean	Median	Minimum	Maximum	Mode	Std Dev	N Miss
V501	2016 GRADE	32873	8.9265963	8.0000000	8.0000000	10.0000000	8.0000000	0.9973175	0
V7221	2016 B01 R HS GRADE/D=1 F1234	31087	6.4878888	7.0000000	1.0000000	9.0000000	9.0000000	2.1505697	1786
V7202	2016 R01 R'S SEX F1234	31300	1.5030351	2.0000000	1.0000000	2.0000000	2.0000000	0.4999988	1573
V7335	2016 B01 LSTYR/U MISBEHAV F1234	32470	1.3871574	1.0000000	1.0000000	5.0000000	1.0000000	0.8148738	403
V7254	2016 M03 TALK PROB W/PRNT F1234	23094	2.0640426	2.0000000	1.0000000	3.0000000	2.0000000	0.7354378	9779
loner		31489	1.8661755	2.0000000	1.0000000	3.0000000	2.0000000	0.6283969	1384

Table 2. Frequencies for categorical variables

Frequencies for categorical variables

The TTEST Procedure

Variable: missdata

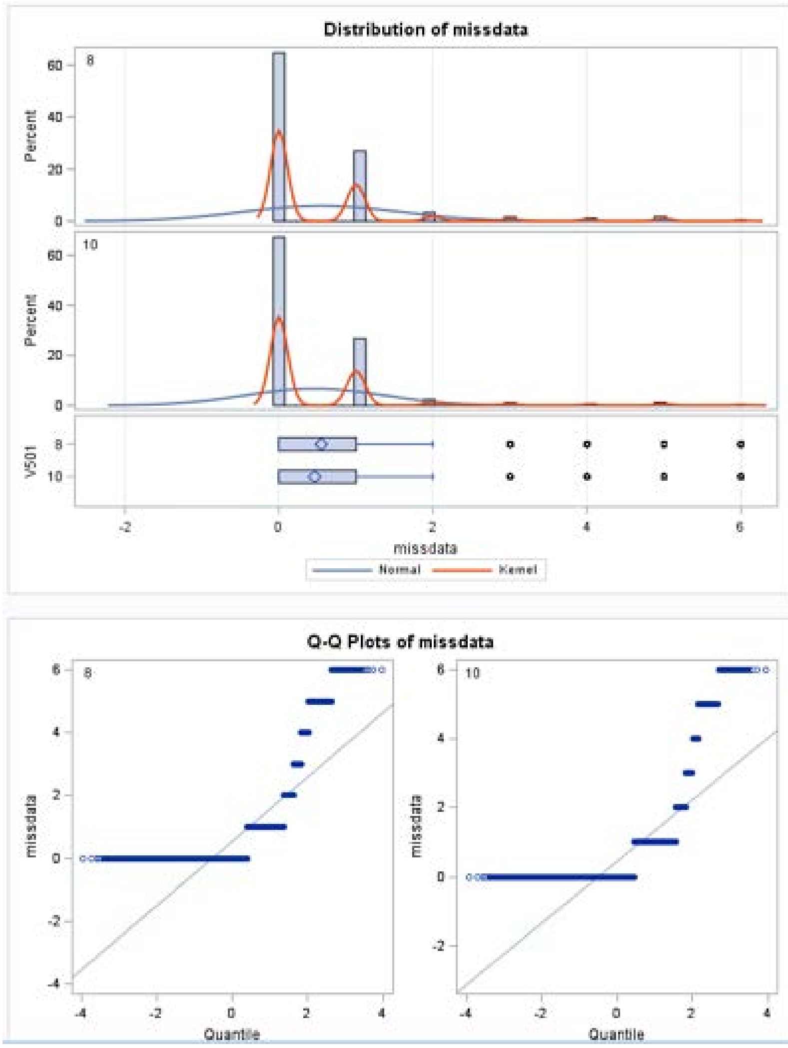
V501	N	Mean	Std Dev	Std Err	Minimum	Maximum
8	17643	0.5495	1.0238	0.00771	0	6.0000
10	15230	0.4607	0.8950	0.00725	0	6.0000
Diff (1-2)		0.0888	0.9662	0.0107		

V501	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
8		0.5495	0.5344	0.5646	1.0238	1.0132	1.0346
10		0.4607	0.4465	0.4749	0.8950	0.8850	0.9051
Diff (1-2)	Pooled	0.0888	0.0679	0.1098	0.9662	0.9589	0.9737
Diff (1-2)	Satterthwaite	0.0888	0.0681	0.1096			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	32871	8.31	<.0001
Satterthwaite	Unequal	32866	8.39	<.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	17642	15229	1.31	<.0001

Figure 1. Distribution of Missing Data (missdata)



STATISTICAL METHODS AND ANALYSIS

SAS 9.4 statistical software was used to conduct a cross-sectional secondary analysis using Pearson’s Chi-Square test statistic for independence to determine the association between “school behavior risk” (v7335), “level of parental communication engaged in” (v7254), the amount of “time spent alone after school” (loner) and student performance (grades). Statistically significant associations were identified between all variables. SAS code for the entire investigation is included in Appendix A.

“School behavior risk” (v7335) and “level of parental communication engaged in” (v7254) had a very strong statistically significant association (Chi-Square = 473.96, $p < .0001$). “school behavior risk” (v7335) and the amount of “time spent alone after school” (loner) also had a strong statistically significant association (Chi-Square = 245.02, $p < .0001$) and finally, “school behavior risk” and “grades” had the strongest statistically significant association (Chi-Square = 2898.16, $p < .0001$).

LOGISTIC REGRESSION

Two logistic regressions were performed using the variables indicated. 32,873 observations were read and 21,986 were used as a result of missing values. This was due to some of the variables not being asked on all of the forms used. Thankfully, the variables selected did contain data for the majority.

For the first model, multinomial logistic regression was selected to consider “behavioral risk” as five discrete categories, rather than a range. “Talking with parents about problems” (v7254), “loner” and “grade” were selected as class variables (Table 4). Values for “grade” are ordered from lowest grade (1 = D) to highest (4 = A). SAS code for the multinomial modeling procedure is shown below:

```
PROC LOGISTIC DATA = mtf DESCENDING plots=oddsratio;
      CLASS v7254 grade loner ;
      MODEL v7335 = v7254 grade loner/ STB rsquare link = glogit ;
RUN ;
```

Table 4: Multinomial modeling detail, convergence, and fit

Class Level Information

Class	Value	Design Variables		
grade	1	1	0	0
	2	0	1	0
	3	0	0	1
	4	-1	-1	-1
V7254	1	1	0	
	2	0	1	
	3	-1	-1	
loner	1	1	0	
	2	0	1	
	3	-1	-1	

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	33943.260	31925.660
SC	33975.253	32181.601
-2 Log L	33935.260	31861.660

Table 5: Multinomial Null Test, Analysis of Effects

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	2073.6000	28	<.0001
Score	2335.9086	28	<.0001
Wald	1922.1285	28	<.0001

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
V7254	8	177.3202	<.0001
loner	8	74.2946	<.0001
grade	12	1486.2888	<.0001

Model output shows that the convergence criterion was satisfied and that fit statistics for the model with covariates vs the mean alone were much better (AIC w/covariates = 31,925.66 < intercept only AIC = 33943.26). The global null test (Table 5) scored very high and shows statistically significant results (Chi-Square = 2073.60, $p < .0001$), thus we reject the null hypothesis that betas = 0.

Our type 3 analysis of effects (Table 5) provide interesting results in that all effects are significant, with “grade” (Chi-Square = 1486.29, $p < .0001$) and “talking with parents about problems” (v7254) (Chi-Square = 177.32, $p < .0001$) having strong effects on the model while "the amount of time spent alone after school" ("loner") (Chi-Square = 74.29, $p < .0001$) did not. This small effect of the "loner" variable was not expected based on our previously run bivariate analysis. This is possibly due to covariance with the other predictors, for example v7254 (Talking with parents about problems).

The Analysis of Maximum Likelihood Estimates (Table 6) showed that all grade effects were statistically significant confirming our type 3 analysis of effects findings. "Talking with parents about problems" (v7254) and "Amount of time spent alone after school" (loner) did have an effect, but only some were statistically significant and not great predictors when compared with "grade."

Table 6: Analysis of Maximum Likelihood Estimates

Analysis of Maximum Likelihood Estimates								
Parameter		V7335	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate
Intercept		5	1	-3.5244	0.0776	2063.5374	<.0001	
Intercept		4	1	-3.0530	0.0655	2174.3108	<.0001	
Intercept		3	1	-2.2156	0.0481	2119.1998	<.0001	
Intercept		2	1	-1.3920	0.0385	1304.0588	<.0001	
V7254	1	5	1	0.5702	0.0866	43.3884	<.0001	0.2307
V7254	1	4	1	0.5597	0.0678	68.1345	<.0001	0.2265
V7254	1	3	1	0.3056	0.0444	47.2937	<.0001	0.1236
V7254	1	2	1	0.1706	0.0305	31.3626	<.0001	0.0690
V7254	2	5	1	-0.4427	0.0950	21.7321	<.0001	-0.2098
V7254	2	4	1	-0.1031	0.0686	2.2584	0.1329	-0.0488
V7254	2	3	1	-0.1042	0.0418	6.2313	0.0126	-0.0494
V7254	2	2	1	0.0345	0.0265	1.6988	0.1925	0.0163
loner	1	5	1	-0.3698	0.1084	11.6429	0.0006	-0.1259
loner	1	4	1	-0.1987	0.0763	6.7737	0.0093	-0.0676
loner	1	3	1	-0.1738	0.0509	11.6719	0.0006	-0.0592
loner	1	2	1	-0.1745	0.0334	27.3314	<.0001	-0.0594
loner	2	5	1	-0.0388	0.0859	0.2039	0.6516	-0.0154
loner	2	4	1	-0.2196	0.0642	11.7134	0.0006	-0.0872
loner	2	3	1	-0.0140	0.0420	0.1115	0.7385	-0.00558
loner	2	2	1	0.0312	0.0276	1.2829	0.2574	0.0124
grade	1	5	1	1.8706	0.1444	167.8292	<.0001	0.5417
grade	1	4	1	1.4260	0.1383	106.3784	<.0001	0.4130
grade	1	3	1	1.0684	0.1148	86.5413	<.0001	0.3094
grade	1	2	1	0.5892	0.1004	34.4417	<.0001	0.1707
grade	2	5	1	0.3556	0.1048	11.5073	0.0007	0.1389
grade	2	4	1	0.6748	0.0813	68.8328	<.0001	0.2637
grade	2	3	1	0.6378	0.0593	115.5977	<.0001	0.2491
grade	2	2	1	0.4366	0.0474	84.9637	<.0001	0.1706
grade	3	5	1	-0.7702	0.1032	55.6874	<.0001	-0.3859
grade	3	4	1	-0.5569	0.0816	46.6345	<.0001	-0.2791
grade	3	3	1	-0.3119	0.0555	31.6304	<.0001	-0.1563
grade	3	2	1	-0.0904	0.0418	4.6759	0.0306	-0.0453

Odds ratio estimates for the model (Figure 3, Table 7) provide details about just how impactful the effects for each coefficient are on our dependent variable outcome. Results were most significant in comparisons between "grades" and "behavioral risk". Output shows that students with lower grades presented a greater "behavioral risk" (v7254) than those with higher grades. For example, compared to students with A's, students with D's were almost 30 X more likely to "always" (OR = 29.73, 95% CI = 18.92 – 46.72), more than 20 X more likely to be in the office "often" (OR = 20.70, 95% CI = 13.6 – 31.54) and about 12 X more likely to be in the office "sometimes" (OR = 11.96, 95% CI = 8.64 – 16.55) due to their "behavioral risk" (v7335). The next most significant contrast was between students with A's versus those with C's who are 9.4 X more likely to be in the office due to behavioral issues "sometimes" (OR = 9.35, 95% CI = 7.03 – 12.44) followed by students with B's who are 2.7 X more likely than students with A's to be in the office "sometimes" (OR = 2.70, 95% CI = 2.03 – 3.59).

Regarding the effects of parental communication (v7254) and "behavioral risk", students who "never" speak to parents about problems were 2.8 X more likely to be in the office "often" compared to student's who spoke with their parents "always" (OR = 2.76, 95% CI = 1.47 – 2.75) and 1.7 X more likely to be in the office "sometimes" (OR = 1.66, 95% CI = 1.41 – 1.95). Interestingly, the least effective predictor of "behavioral risk" (v7335) is the amount of "time students spend alone after school" (loner). Results showed that there is a small increased risk for students spending more time alone after school, but the effect is minimal. For this reason, odds ratios will not be discussed for this variable, but are made available in the table below.

The visualization below (Figure 3) provides important insights into the value of multinomial regression as a "detail" or "in-group" level utility. It easy to visually compare sub-populations of students against one and other. For our case study this multinomial approach is useful because, we are better able to profile the most likely student candidates for behavioral interventions. Further, this level of detail more accurately represents the spectrum of behavioral needs known to exist and would be useful in designing and deploying future intervention programs.

Figure 3: Multinomial Odds Ratio Visualization (class detail)

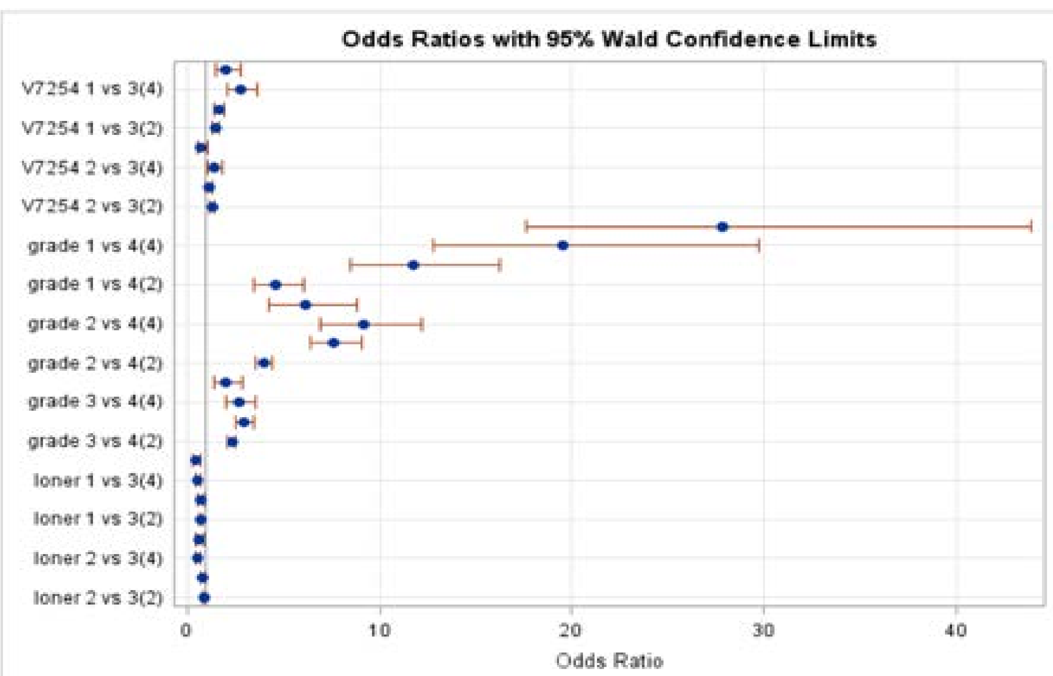


Table 7: Multinomial Odds Ratio Estimates

Odds Ratio Estimates				
Effect	V7335	Point Estimate	95% Wald Confidence Limits	
V7254 1 vs 3	5	2.009	1.466	2.752
V7254 1 vs 3	4	2.763	2.111	3.617
V7254 1 vs 3	3	1.660	1.412	1.951
V7254 1 vs 3	2	1.456	1.306	1.622
V7254 2 vs 3	5	0.730	0.519	1.027
V7254 2 vs 3	4	1.424	1.086	1.868
V7254 2 vs 3	3	1.102	0.945	1.285
V7254 2 vs 3	2	1.271	1.155	1.398
loner 1 vs 3	5	0.459	0.314	0.672
loner 1 vs 3	4	0.540	0.411	0.708
loner 1 vs 3	3	0.696	0.575	0.844
loner 1 vs 3	2	0.728	0.640	0.828
loner 2 vs 3	5	0.639	0.470	0.870
loner 2 vs 3	4	0.528	0.419	0.666
loner 2 vs 3	3	0.817	0.693	0.964
loner 2 vs 3	2	0.894	0.800	1.000
grade 1 vs 4	5	27.834	17.656	43.880
grade 1 vs 4	4	19.489	12.783	29.711
grade 1 vs 4	3	11.733	8.470	16.254
grade 1 vs 4	2	4.594	3.506	6.018
grade 2 vs 4	5	6.119	4.262	8.784
grade 2 vs 4	4	9.196	6.913	12.232
grade 2 vs 4	3	7.627	6.382	9.115
grade 2 vs 4	2	3.944	3.524	4.414
grade 3 vs 4	5	1.985	1.391	2.832
grade 3 vs 4	4	2.683	2.019	3.565
grade 3 vs 4	3	2.951	2.497	3.488
grade 3 vs 4	2	2.328	2.121	2.556

For our second model, Ordinal Regression was invoked to contrast the impact of model selection on output and visualization (Figure 4). The SAS code used to produce our ordinal model is shown below:

```
PROC LOGISTIC DATA = mtfbin DESCENDING plots=oddsratio;
      CLASS v7254 grade loner ;
      MODEL v7335 = v7254 grade loner/ STB rsquare ;
RUN ;
```

Table 8 displays the convergence criteria for our ordinal model, which is again satisfied. A comparison of fit statistics shows about 1% decline compared to the Multinomial Model (Ordinal Model AIC = 32001.79 > Multinomial Model AIC = 31925.66). The adjusted R^2 indicates that this model explains 8.5% of the variance (R-square = 0.085). These tell us right away that both models fit our data equally, so which do we choose? Let us compare each model.

Table 8: Ordinal Model Convergence, Fit, R^2

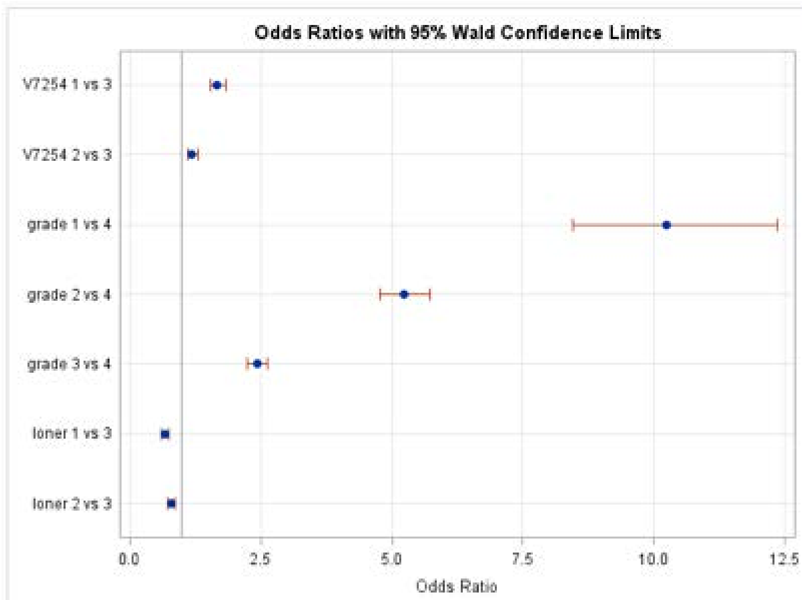
Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Score Test for the Proportional Odds Assumption		
Chi-Square	DF	Pr > ChiSq
133.9258	21	<.0001

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	33943.260	32001.786
SC	33975.253	32089.765
-2 Log L	33935.260	31979.786

R-Square	0.0851	Max-rescaled R-Square	0.1082

Figure 4: Ordinal Model Output Visualization



The multinomial model appears to present the effects of in-class variables on the dependent variable in more detail. This is useful if we are interested in presenting our results in a way that more accurately represents the spectrum of behavioral needs known to exist rather than as a nested group and would be useful for the purposes of designing and deploying future intervention programs.

Our ordinal model, however, produced results from a much more generalized perspective, a sort of 'high-level' view. This could be useful if a cursory investigation was required, for example, to justify the need for a program in the first place. Perhaps to justify the need for additional funding in support of a more detailed investigation, such as the one our multinomial model provided.

Our global null test (Table 9) again scored very high showing and statistically significant results (Chi-Square = 1958.59, p < .0001), thus we reject the null hypothesis that betas =

0. Our type 3 analysis of effects provided similar results and all effects were again significant and very similar to the Multinomial

Model with "grade" (Chi-Square = 1477.16, $p < .0001$), "loner" (Chi-Square = 58.32, $p < .0001$) and "talking with parents about problems" (v7254) (Chi-Square = 141.04, $p < .0001$).

Table 9: Ordinal Model H0, Analysis of Effects

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	1955.4744	7	<.0001
Score	1958.5900	7	<.0001
Wald	1873.6691	7	<.0001

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
V7254	2	141.0392	<.0001
grade	3	1477.1602	<.0001
loner	2	58.3219	<.0001

Table 10: Ordinal Analysis of Max Likelihood

Analysis of Maximum Likelihood Estimates							
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate
Intercept	5	1	-4.0754	0.0655	3870.9519	<.0001	
Intercept	4	1	-3.0106	0.0430	4896.3897	<.0001	
Intercept	3	1	-1.9407	0.0322	3631.0476	<.0001	
Intercept	2	1	-0.7122	0.0281	643.0601	<.0001	
V7254	1	1	0.2848	0.0245	134.9517	<.0001	0.1152
V7254	2	1	-0.0538	0.0221	5.9088	0.0151	-0.0255
grade	1	1	1.1083	0.0694	255.2651	<.0001	0.3210
grade	2	1	0.4386	0.0351	156.0418	<.0001	0.1714
grade	3	1	-0.3302	0.0313	111.5058	<.0001	-0.1655
loner	1	1	-0.1910	0.0273	48.8690	<.0001	-0.0650
loner	2	1	-0.0229	0.0226	1.0260	0.3111	-0.00911

The ordinal model's analysis of effects results indicate again that "grades" are the coefficient most in agreement with the observed data, and while "Talking with parents about problems" (v7254) and "Amount of time spent alone after school" (loner) did have an effect, both were less effective predictors of behavior than grade. The Analysis of Maximum Likelihood Estimates (Table 10) showed that most effects were statistically significant, with exceptions for students who spent 1 -3 hours alone ("loner" = 2) after school and kids who "spoke to their parents sometimes."

Odds ratio estimates for the model (Table 11) provided details about just how impactful the effects for each coefficient are on our dependent variable outcome.

Table 11: Ordinal Model Odds Ratio Estimates

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
V7254 1 vs 3	1.675	1.534	1.829
V7254 2 vs 3	1.194	1.102	1.294
grade 1 vs 4	10.227	8.457	12.367
grade 2 vs 4	5.235	4.773	5.741
grade 3 vs 4	2.427	2.241	2.628
loner 1 vs 3	0.667	0.601	0.740
loner 2 vs 3	0.789	0.721	0.863

As expected, output shows that the comparison between "grades" and "behavioral risk" were most significant. Students with lower grades presented a greater "behavioral risk" (v7254) than those with higher grades. For example, students with D's, on average, were 10 X more likely to display behavioral issues compared to students with A's (OR = 10.23 95% CI = 8.46 – 12.34), more than 5.0 X more likely than students with B's (OR = 5.24 95% CI = 4.77 – 5.74) and about 2.5 X more likely to display behavioral issues than students with C's (OR = 2.43, 95% CI = 2.24 – 2.63).

Regarding the effects of parental communication (v7254) and "behavioral risk", our ordinal model shows that students who "never" speak to parents about problems were 1.7 X more likely to present a "behavioral risk" compared to student's who spoke with their parents "always" (OR = 1.68, 95% CI = 1.53 – 1.83) and 1.2 X more likely than students who spoke to their parents about "some" problems (OR = 1.21, 95% CI = 1.10 – 1.30).

In agreement with our Multinomial Model, the least effective predictor of "behavioral risk" (v7335) is the amount of "time students spend alone after school" (loner). Results showed that, while there is a small increased risk for students who spend time alone after school, the amount of time does not appear to magnify this effect substantially suggesting more of a binary relationship (i.e. those who spend time alone vs. those who do not).

The association of predicted probabilities and observed responses (Table 12) and associated odds ratio output (Figure 4) indicates 65.3% concordance. Somers' D (0.367) indicates that 38.1% of behavioral issues can be predicted by better grades and having at least "some" communication with parents. The odds ratio estimates are within 95% confidence limits and variable level comparisons mirrored our findings from the Multinomial model thus, they will not be re-examined here.

Another benefit of the ordinal procedure is that visualizations are not restricted to 5000 data-points like the multinomial procedure, thus making visual comparisons of the overall effects of all variables easier by eliminating the need for additional coding. For this investigation

Figure 4: Ordinal Model Output Visualization

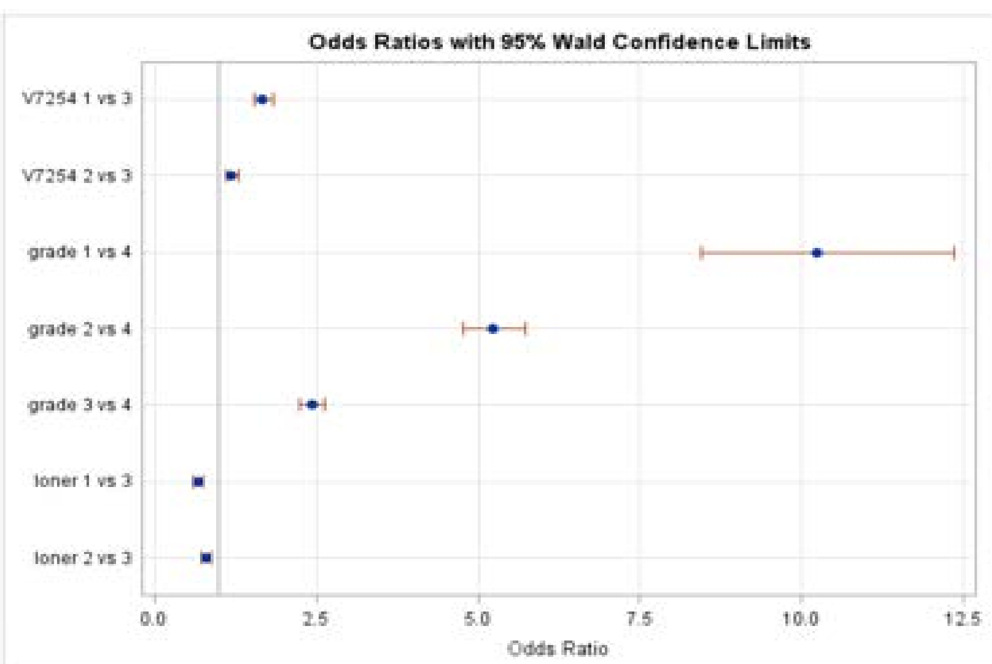


Table 12: Ordinal – Association of Predicted Probabilities and Observed Response

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	65.3	Somers' D	0.367
Percent Discordant	28.5	Gamma	0.392
Percent Tied	6.2	Tau-a	0.141
Pairs	92762571	c	0.684

specifically, visualizations produced from our ordinal model (Figure 4) plainly show the greatest difference in "behavioral risk" existing between students who receive A's and D's followed by those who receive C's and A's and finally those receiving B's and A's.

To provide a fair comparison of both models benefits we note that output form our:

- multinomial model (Figure 2)
    - provides a greater level of detail, especially where the DV is a spectrum
    - maximizes in-class contrast
    - requires additional SAS code when > 5000 data points exist
- ordinal model (Figure 4)
    - better for generalizing findings
    - limited in-class contrast
    - provides a better vector for "high-level" or top-down comparisons
    - does not require additional coding

While both models performed very well, user intent (i.e. intended use for data) is the ultimate deciding factor in selection.

CONCLUSION

This has been a review of two logistic regression models to explore the predictive factors associated with "behavioral risk" of students whom may require a range of intervention measures. While additional research should be conducted to determine root causes of such behavioral risks, the data appear to support the notion that grades are the best predictor of "behavioral risk" and students whom talk with parents about some or more problems tend to get better grades. In other words, if you want your kids to get better grades and stay out of trouble, talk to them.

Additionally, this review explored key differences between multinomial and ordinal regression model outputs. We discovered that multinomial models provide an advantage over ordinal models in visualizing in-class variable effects on the dependent variable(s). These visualizations, however, can be difficult to interpret and require additional coding in SAS when more than 5000 data-points exist. We also discovered that ordinal models are better for producing generalizing visualizations to compare independent variables effects from a class-level perspective. That is, in the context of the overall model (i.e. a top-down perspective).

Researchers and program administrators should be mindful of these differences when selecting a representative model. It is easy to imagine a scenario in which model selection fails to appropriately convey the intention. In our case study, a behavioral intervention program intended to be designed in a general capacity (i.e. to support all students), should be presented with ordinal output, whereas, downstream intervention program development will be better supported with the multinomial model's additional level of in-class detail. For example, to support varying levels of intervention for students with the most severe behavior problems.

REFERENCES

Hosmer DW Jr, Lemeshow S, Sturdivant RX (2013) Applied logistic regression, vol 398. Wiley, New York

Johnston, et. al. Monitoring the Future: A Continuing Study of American Youth (8th- and 10th-Grade Surveys), 2016. Inter-University Consortium for Political and Social Research (ICPSR 36799). www.icpsr.umich.edu

Wayland, Jon. What are the most in demand skills for Data Scientists in 2019? Quora. Published 01/23/2019. Last accessed 07/01/2019 https://www.quora.com/What-are-the-most-in-demand-skills-for-Data-Scientists-in-2019.

ACKNOWLEDGMENTS

Dr. Tyler Smith – National University

Dr. AnnMaria De Mars – National University, The Julia Group, 7th Generation Games

Kirk Lafler, M.S. – Software Intelligence Corporation, SAS Author/Consultant

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Corey Leadbeater, M.S. Applied Data Science
8People, Inc. and National University
(619) 928 - 1221
corey.leadbeater@8peopleinc.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.

APPENDIX A: SAS CODE

```
Data MTF;
   set Futures.Da36799p1 ;
/*Demographic Info*/
      if v7221 < 1 then v7221 = . ; /* Avg Grades Received */
      if v7202 < 1 then v7202 = . ; /* gender: 1 = m, 2 = f */
/*Behavior*/
      if v7335 < 1 then v7335 = . ; /*Student Behavior at school, 1 = never,
2 = seldom, 3 = sometimes, 4 = often, 5 = always*/
/*Parental Involvement, Structure */
      if v7254 < 1 then v7254 = . ; /* talk with parents about problems - 1 =
no, 2 = some, 3 = lots */
/*Time alone*/
      if v7214 < 1 then v7214 = . ; /*time spent along after school, 1 =
none, 2 = 1hr, 3 = 1-2 hrs, 4 = 2-3 hrs, 5 = 4-5hrs, 6 = >5hrs */
/*recode of grades into smaller categories*/
      if v7221 in (1) then grade = 1 ; /* Grade = D */
      if v7221 in (2,3,4) then grade = 2 ; /* Grade = C */
      if v7221 in (5,6,7) then grade = 3 ; /* Grade = B */
      if v7221 in (8,9) then grade = 4 ; /* Grade = A */
/*recode time alone to categories*/
      if v7214 in (1) then loner = 1 ; /*no time alone */
      if v7214 in (2,3,4) then loner = 2 ; /*1-3 hrs alone*/
      if v7214 in (5,6) then loner = 3 ; /*4 or more hrs alone*/
  keep v501 v7221 v7202 v7335 v7254 v7214 loner grade; /*v501 = grade level 8
or 10 */
 run;
Title " Descriptive statistics for numeric variables" ;
 proc means data=Mtf n mean median min max mode stddev nmiss;
  var  v501 v7221 v7202 v7335 v7254 loner ;
 run ;
 Title " Descriptive statistics for numeric variables" ;
 proc univariate data=Mtf ;
  var  v501 v7221 v7202 v7335 v7254 loner ;
 run ;
 Title "Frequencies for categorical variables" ;
 proc freq data=mtf ;
   tables  v501 v7221 v7202 v7335 v7254 loner grade;
   run;
*** Missing data check ;
Data check ;
  set mtf ;
  array chckmiss{*} v501 v7221 v7202 v7335 v7254 loner grade ;
  missdata = 0 ;
  do i = 1 to dim(chckmiss) ; /* dim(chckmiss) = set dimensions
  to however variables are in chckmiss */
     if chckmiss{i} = . then missdata = missdata + 1 ;
  end ;
  if missdata > 0 then anymiss = 1 ;
     else anymiss = 0 ;
run;
*** Check for missing data differences by grade ;
proc freq data =check ;
  tables anymiss missdata ;
  run ;
proc freq data=check ;
  tables   anymiss*(v501 v7221 v7202 v7335 v7254 loner grade) ;
```

```
  run ;
  proc ttest data=check ;
  class v501 ;
    var missdata ;
    run;
/*Cross Tabulation with chi-square of cat variable predictors*/
TITLE "Cross Tabulation with cat variable predictors";
Proc freq data=mtf;
      tables v7335*(v7254 loner grade) / chisq; /*v7335 = stu behavior at
school,
      v7254 = talk to parents about probs, loner = time alone each day after
school */
      run;
TITLE "ANOVA 1: 1 numeric independent variable, 1 dependent - tukey post hoc"
;
proc glm data=mtf plots = all;
   class v7254 ;
   model v7335 = v7254; /* student behavior at school (v7335) = talk to
parents about problems (v7254)*/
   MEANS v7254/ tukey ;
run;
TITLE "Multinomial Logistic Regression: 3 independent variables, 1 dependent
variable" ;
PROC LOGISTIC DATA = mtf DESCENDING plots=oddsratio;
      CLASS v7254 grade loner ;
      MODEL v7335 = v7254 grade loner/ STB rsquare link = glogit ; /*v7335
run as multinomial with "link = glogit" since there are varying levels of
behavioral risk. These varying levels of trouble assists us in determining
what level of intervention is most appropriate for students */
RUN ;
TITLE "Ordinal Logistic Regression: 3 independent variables, 1 dependent
variable" ;
PROC LOGISTIC DATA = mtfbin DESCENDING plots=oddsratio;
      CLASS v7254 grade loner ;
      MODEL v7335 = v7254 grade loner/ STB rsquare ;
RUN ;
```